

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.**

(19) 日本国特許庁 (J P)

(12) 公開特許公報 (A)

(11) 特許出願公開番号
特開2000-76300
(P2000-76300A)

(43) 公開日 平成12年3月14日 (2000.3.14)

(51) Int.Cl.⁷
G 0 6 F 17/30

識別記号

F I
G 0 6 F 15/40
15/401

テーマコード(参考)

3 7 0 G
3 1 0 C
3 3 0 Z

審査請求 未請求 請求項の数 2 O L (全 9 頁)

(21) 出願番号 特願平10-249766
(22) 出願日 平成10年9月3日 (1998.9.3)

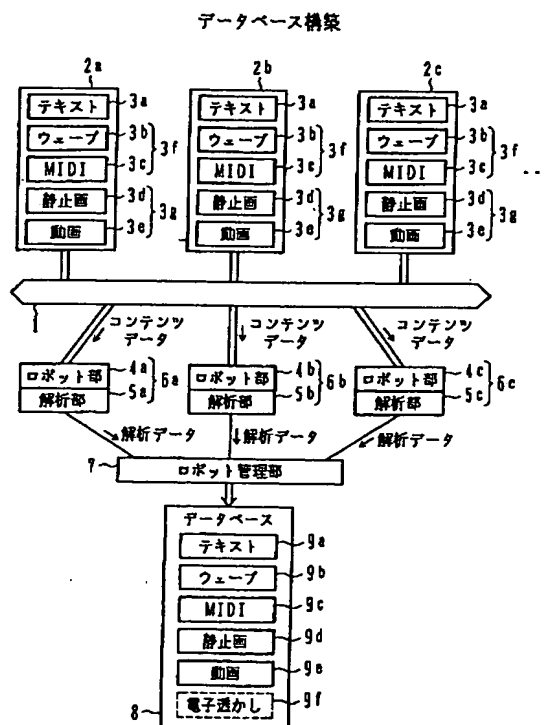
(71) 出願人 000004075
ヤマハ株式会社
静岡県浜松市中沢町10番1号
(72) 発明者 樽口 秀昭
静岡県浜松市中沢町10番1号 ヤマハ株式
会社内
(74) 代理人 100091340
弁理士 高橋 敬四郎 (外1名)

(54) 【発明の名称】 データベース構築装置及び検索エンジン

(57) 【要約】

【課題】 テキストデータ以外のメディアデータを基にデータベースを構築することができるデータベース構築装置及びそのデータベースを用いた検索エンジンを提供することを課題とする。

【解決手段】 ネットワーク上のサイトからメディアデータを読み出す読み出し手段(4)と、読み出したメディアデータのうちテキストデータ以外のメディアデータを解析して鍵を検出する解析手段(5)と、前記鍵と前記サイトとを対応付けてデータベース(8)を構築する構築手段(7)とを有する検索エンジン用のデータベース構築装置。



【特許請求の範囲】

【請求項1】 ネットワーク上のサイトからメディアデータを読み出す読み出し手段と、

前記読み出したメディアデータのうちテキストデータ以外のメディアデータを解析して鍵を検出する解析手段と、

前記鍵と前記サイトとを対応付けてデータベースを構築する構築手段とを有する検索エンジン用のデータベース構築装置。

【請求項2】 テキストデータ以外のメディアデータに関連する鍵とそのデータが存在するネットワーク上のサイトとを対応付けたデータベースと、

検索のための鍵を指定する指定手段と、

前記データベースを基に前記指定された鍵に関連するテキストデータ以外のメディアデータが存在するネットワーク上のサイトを検索する検索手段とを有する検索エンジン。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】 本発明は、データ検索技術に関し、特にネットワーク上のデータを基にしてデータベースを構築するデータベース構築装置及びそのデータベースを用いた検索エンジンに関する。

【0002】

【従来の技術】 現在、インターネットが普及している。ユーザは、インターネットに接続すれば、膨大な数のホームページを見ることができる。多数のホームページの中から、所望のホームページを探すことは困難である。検索エンジンを用いれば、所望のホームページを検索することができる。

【0003】 検索エンジンは、データベースに接続され（又はデータベースを有し）、一定の条件に該当するホームページの一覧表示を行う。検索エンジンには、ロボット型とディレクトリ型の2種類がある。

【0004】 ロボット型検索エンジンは、世界中のサーバを定期的に自動巡回して収集したホームページの情報をデータベース化する。自動巡回するので、データベースに登録されるホームページの数が多いのが利点である。データベースは、ホームページの中で使用されているテキストデータ（文字データ）を基に生成される。

【0005】 ホームページの検索は、ユーザが指定したテキストデータをキーワードとして行われる。キーワードが適切でない場合には、見当外れのホームページが検索されてしまう。

【0006】 ディレクトリ型検索エンジンは、ホームページの作者又は検索エンジンの管理者が登録することにより、データベースが構築される。特定者が登録を行うので、データベースに登録されるホームページの数は少ないが、その信頼性は高い。すなわち、データベースは分類分けされているので、無関係なホームページが検索

される可能性は低い。ユーザは、キーワードを指定することにより、所望のホームページを検索することができる。

【0007】

【発明が解決しようとする課題】 従来は、テキストデータの通信が主であったが、近年、マルチメディアの通信が増加してきている。例えば、テキストデータの他、音楽データや画像データの通信が行われる。ホームページにおいても、テキストデータのみならず、音楽データや画像データが使用されている。

【0008】 しかし、検索エンジンは、テキストデータを検索することは可能であるが、所望の音楽データや画像データを検索することが困難である。テキストデータ以外のメディアデータを効率よく検索することができる検索エンジンが望まれている。

【0009】 本発明の目的は、テキストデータ以外のメディアデータを基にデータベースを構築することができるデータベース構築装置を提供することである。

【0010】 本発明の他の目的は、テキストデータ以外のメディアデータを基に構築されたデータベースを用いて検索を行う検索エンジンを提供することである。

【0011】

【課題を解決するための手段】 本発明の一観点によれば、ネットワーク上のサイトからメディアデータを読み出す読み出し手段と、前記読み出したメディアデータのうちテキストデータ以外のメディアデータを解析して鍵を検出する解析手段と、前記鍵と前記サイトとを対応付けてデータベースを構築する構築手段とを有する検索エンジン用のデータベース構築装置が提供される。

【0012】 本発明の他の観点によれば、テキストデータ以外のメディアデータに関連する鍵とそのデータが存在するネットワーク上のサイトとを対応付けたデータベースと、検索のための鍵を指定する指定手段と、前記データベースを基に前記指定された鍵に関連するテキストデータ以外のメディアデータが存在するネットワーク上のサイトを検索する検索手段とを有する検索エンジンが提供される。

【0013】 データベースには、テキストデータ以外のメディアデータに関連する鍵とそのデータが存在するネットワーク上のサイトとが対応付けて登録される。ユーザは、所定の鍵を指定することにより、その鍵に関連するテキストデータ以外のメディアデータが存在するネットワーク上のサイトを検索することができる。

【0014】

【発明の実施の形態】 図1は、本発明の実施例によるデータベース構築装置の構成を示すブロック図である。

【0015】 インターネット回線1には、多数のサイト（ホームページ）2a、2b、2c、・・・が接続されている。以下、サイト2a、2b、2cの個々又は全てをサイト2という。

【0016】各サイト2は、メディアデータを有する。例えば、テキストデータ3a、楽音のウェーブデータ3b、MIDI (musical instrument digital interface) データ3c、静止画データ3d及び動画データ3eの中の1つ又は任意の組み合わせを有する。

【0017】上記のデータは、大きく分けると、テキストデータ3a、楽音データ3f (ウェーブデータ3b及びMIDIデータ3cを含む)、及び画像データ3g (静止画データ3d及び動画データ3eを含む) に分類される。

【0018】テキストデータ3aは、文字データであり、例えばHTML (hyper text markup language) 形式で記述されたテキストデータである。インターネット上にホームページを開設する場合には、HTML形式で記述すれば、関連する情報にリンクすることができ、さらに、画面に表示するイメージファイルの名前、又はリンク先のサーバ名とファイル名を指定することもできる。

【0019】ウェーブデータ3bは、例えばマイクで楽音を電気信号に変換し、その電気信号をアナログ形式からデジタル形式に変換したデータである。ウェーブデータ3bをアナログ形式に変換して、スピーカに供給すれば、楽音を発音させることができる。ウェーブデータ3bは、楽器音及び音声のデータを含む。

【0020】MIDIデータ3cは、複数の電子楽器間を接続するためのMIDI規格に準拠したデータであり、ノートオンイベント及びノートオフイベント等を含む。ノートオンイベントは、発音開始を指示するイベントであり、例えば押鍵操作により生成される。ノートオフイベントは、消音を指示するイベントであり、例えば離鍵操作により生成される。

【0021】静止画データ3dは、例えばJPEG (joint photographic experts group) 規格により圧縮された静止画データである。現在、パーソナルコンピュータやワークステーションのほとんどが、静止画の圧縮方式としてJPEG圧縮方式を採用している。例えば、静止画データ3dは、デジタルカメラにより生成される。

【0022】動画データ3eは、例えばMPEG (moving picture experts group) 規格により圧縮された動画データである。MPEGは、さらにMPEG1、MPEG2及びMPEG4に分けることができる。

【0023】データベース構築装置は、巡回及び解析部6a、6b、6c、ロボット管理部7、及びデータベース部8を有する。巡回及び解析部6aは、巡回ロボット部4a及びコンテンツ解析部5aを有する。同様に、巡回及び解析部6b及び6cは、それぞれ巡回ロボット部4b、4c及びコンテンツ解析部5b、5cを有する。巡回及び解析部6a、6b、6cは、3個に限定されず、それより少なくとも多くてもよい。

【0024】以下、巡回及び解析部6a、6b、6cの

個々又は全てを、巡回及び解析部6といい、巡回ロボット部4a、4b、4cの個々又は全てを、巡回ロボット部4といい、コンテンツ解析部5a、5b、5cの個々又は全てを、コンテンツ解析部5という。

【0025】巡回ロボット部4は、インターネット回線1を介して、可能な限り多くのサイト2を定期的に自動巡回し、サイト2内のコンテンツを読み出す (ダウンロードする)。すなわち、このデータベース構築装置は、ロボット型である。コンテンツは、例えばテキストデータ3a、ウェーブデータ3b、MIDIデータ3c、静止画データ3d、及び/又は動画データ3eを含む。具体的には、巡回ロボット部4は、各サイト2のアドレスを示すURL (uniform resource locator) を指定し、そのサイト2内のコンテンツを読み出す。

【0026】コンテンツ解析部5は、巡回ロボット部4が読み出したコンテンツの種類を分類する。次に、その分類方法を説明する。

【0027】図2は、コンテンツの分類方法を示すブロック図である。サイト2には、様々な種類のメディアデータが含まれている。各メディアデータは、例えばテキストデータ3a、ウェーブデータ3b、MIDIデータ3c、静止画データ3d又は動画データ3eのいずれかに分類される。

【0028】具体的な分類方法の例を3つ示す。

(1) ファイル名の拡張子を判定する方法

テキストデータ3a、ウェーブデータ3b、MIDIデータ3c等は、それぞれ別のファイルに収められている。各ファイルには、ファイル名が付与されている。例えば、ファイル名は、「ABCD. html」で表される。ファイル名の後半部の「html」が拡張子である。

【0029】拡張子が「html」又は「htm」の時には、HTML形式で記述されたテキストデータ3aであると判断することができる。

【0030】拡張子が「wav」である時には、ウェーブデータ3bであると判断することができる。

【0031】拡張子が「mid」である時には、MIDIデータ3cであると判断することができる。画像データ3d、3eについても拡張子を基に分類することができる。

【0032】(2) ファイル内のチャンクデータを判定する方法

ファイル内には、チャンクデータが含まれているものがある。ファイルを実際に読み込み、そのチャンクデータを基にコンテンツの分類を行うことができる。

【0033】(3) MIME-TYPEを取得する方法
MIME (multipurpose internet mail extension) -TYPEは、コンテンツのタイプを示す情報であり、HTTP (hyper-text transfer protocol) で取得することができる。HTTPは、インターネット通信で使

れるプロトコルである。

【0034】MIME-TYPEは、HTML形式のテキストデータ3aの場合は「text/html」、ウェブデータ3bの場合は「audio/wav」、MIDIデータ3cの場合は「audio/mid」又は「smf」で表される。smfは、MIDIデータの標準ファイル形式であるスタンダードMIDIファイルを意味する。

【0035】図1において、データベース8には、テキストデータ9a、ウェブデータ9b、MIDIデータ9c、静止画データ9d、及び動画データ9eの他に、電子透かしデータ9fの各情報が登録される。次に、電子透かしデータ9fについて説明する。

【0036】電子透かしデータは、インターネット上でのコンテンツの不正コピーを防止するために用いられる。コンテンツ内には、著作権情報が電子透かしデータとして埋め込まれる。この電子透かしデータをコンテンツ内から検出し、不正の有無を判定することができる。

【0037】電子透かしデータは、コンテンツ内に巧妙に埋め込まれ、第三者には検出することができない。例えば、電子透かし技術の一例は、本出願人の別の出願である特願平9-193128号の図1及びその説明部分に記載されている。また、特願平8-191528号及び特願平9-148612号にも記載されている。これらの出願には、MIDIデータのノートオンイベント内のベロシティに電子透かしデータを埋め込む方法が記載されている。

【0038】電子透かしデータは、MIDIデータ3cの他、ウェブデータ3b、静止画データ3d及び動画データ3eに埋め込むこともできる。

【0039】図1に示すコンテンツ解析部5は、コンテンツを解析し、MIDIデータ3c等から電子透かしデータを検出することができる。

【0040】コンテンツ解析部5は、分類されたコンテンツ毎に、以下のコンテンツの内容を解析する。

【0041】(1) MIDIデータ

MIDIデータは、プログラムチェンジ、キーイベント及びメタイイベントを含む。プログラムチェンジは音色を示すので、プログラムチェンジを解析することにより、音色をデータベースの登録項目にすることができる。

【0042】キーイベントは、発音開始を表すキーオンイベントと消音を表すキーオフイベントに分けることができる。キーイベントは、キーコード（音高）及びベロシティ（音量）を含む。キーコード及びベロシティを登録項目にする他、前のキーとの差分値、全てのキーについての平均値、分散、最大値又は最小値を登録項目にすることができる。

【0043】メタイイベントは、テンポ、拍子、曲調、シークセンサの種類等を表すので、メタイイベントを解析することにより、それらを登録項目にすることができる。

【0044】典型的フレーズを検出して、それを登録項目にすることができる。典型的フレーズは、いわゆる「さび」に相当し、高い頻度で繰り返し現れるフレーズ又は類似したパターンが高い頻度で表れるフレーズである。典型的フレーズを登録すれば、ユーザはそのフレーズを鍵（キー）として検索することが可能になる。後に説明する検索では、入力されたフレーズ又はそれに近いものを検索して、リストアップすることができる。

【0045】曲構成を解析し、登録項目にする。例えば、ブロック（○拍、○小節）のパターン分類後、ブロックによる曲構成情報を登録項目にする。

【0046】曲又はフレーズの時間推移を解析し、登録項目にする。例えば、ブロック間の変化情報を差分値等として求める。また、音色の変化を解析することもできる。

【0047】統計的解析を行い、登録項目にする。例えば、音符又は休符の時間、音高又は音程の平均値や分散値、同時発音の音の組み合わせを解析する。

【0048】上記の情報から作曲者又は演奏者のくせを見抜くことが可能になり、イメージ的に目的のMIDIデータを検索することができる。また、音楽のジャンルを解析することもできる。

【0049】(2) ウェブデータ

音声認識、周波数解析、時間推移解析を行う。例えば、男性又は女性、音声認識された語句（単語）、楽器の種類等を解析することができる。

【0050】(3) 静止画データ

静止画の間引きを行って画素数を減らした概画像を生成したり、画素値の統計的解析を行ったり、周波数解析を行う。また、静止画の配色パターンを解析することもできる。

【0051】(4) 動画データ

動画のカット部分の静止画を解析したり、動画の時間推移解析を行う。

【0052】(5) その他

テキストデータ、電子透かしデータ、ファイルのヘッダ情報、及びメタテキストの情報については、以下の項目を解析する。

【0053】曲名、アルバム名、作曲者、編曲者、作詞者、演奏者、歌手、版元、著作権者、制作者、技術者、使用機器、スタジオ、楽器、制作日時、関連アドレス（地理的アドレス及びネットワーク的アドレスを含む）、関連リンク情報、値段、使用条件（環境、使用者、期限）等を解析する。これらの情報は、テキストデータ又はコードで表すことができる。コードで表す場合は、そのコードと対応する情報を結び付けるためのテーブルを用意する必要がある。

【0054】データベース8には、テキストデータ9aの他、楽音データ9b、9c、画像データ9d、9e、及び電子透かしデータ9fの情報が登録されている。ユ

ーザは、テキストデータを鍵として楽音データや画像データを検索することができる。

【0055】例えば、曲名（テキストデータ）を鍵として、その曲名のMIDIデータやウェブデータを検索したり、著作権者（テキストデータ）を鍵として、画像データ（絵画や写真）を検索することができる。また、著作権情報が電子透かしデータとして存在する場合には、電子透かしデータを解析し、著作権管理を行うことができる。

【0056】また、楽音データを鍵として楽音データを検索したり、画像データを鍵として画像データを検索することもできる。

【0057】本実施例によれば、テキストデータ以外のメディアデータをデータベース化し、検索対象にすることができる。テキストデータ以外のメディアデータとしては、例えば、MIDIデータ、ウェブデータ、静止画データ、動画データ等であり、特にMIDIデータが好ましい。また、テキストデータ以外のメディアデータを解析し、その中に存在する電子透かしデータの情報を検索のための鍵として検出することができる。

【0058】次に、図1に示すロボット管理部7について説明する。ロボット管理部7は、コンテンツ解析部5から供給された解析データを受け、その解析データを鍵として対応するサイト2及びそのURLをデータベース8に登録する。データベース8には、どんなコンテンツがどのサイト（URL）2に存在するのかの情報を蓄積する。

【0059】データベース8は、テキストデータ9aのみではなく、ウェブデータ9b、MIDIデータ9c、静止画データ9d、動画データ9e及び電子透かしデータ9fに関する情報（鍵）が登録される。

【0060】例えば、「女性の音声」を鍵としてウェブデータ9bを検索、又は「バイオリンの音色」を鍵としてMIDIデータ9cを検索することができる。

【0061】データベース8には、上記の解析データの他、データの種類、データの大きさ（又は再生時間）、再生ソフトの種類（楽音再生ソフト又は画像再生ソフト）、リンク先、データベースの構築日時、パスワードの有無等を登録することができる。これらの項目は、検索の際に画面に表示して、ユーザに知らせることができる。

【0062】図3は、上記のデータベースを用いて検索する方法を示す。上記で構築されたデータベース8は、インターネット回線1に接続される。ユーザは、検索エンジン11をインターネット回線1に接続することにより、所望のメディアデータの検索を行うことができる。

【0063】ユーザは、MIDIデータ又は静止画データ等のデータ種類を指定することにより、そのデータ種類のデータを有するサイトを検索することができる。サイトにはそのアドレス（URL）が対応付けられている

ので、そのサイトを選択すれば、そのサイトのホームページを見ることができる。

【0064】また、ユーザは、「女性の声」等のテキストデータを鍵として指定して、ウェブデータを検索することができる。また、楽音や画像を鍵として、楽音データや画像データを検索することもできる。例えば、MIDIデータや楽譜（音符）データを鍵として指定することもできる。

【0065】ユーザは、種々のメディアデータを検索の鍵として入力することができる。例えば、キーボードやマウスを使用することにより、テキストデータを鍵として入力することができる。また、画面に表示された画像又はファイル内の画像の全部又は一部を切り取って、画像を鍵として入力することができる。また、マイクを使ってウェブデータを入力したり、楽譜入力装置やシーケンサを使ってMIDIデータを入力することができる。

【0066】データベース8には、テキストデータ9a、ウェブデータ9b、MIDIデータ9c、静止画データ9d、動画データ9e及び／又は電子透かしデータ9fについての情報が登録されている。

【0067】データベース8に登録された情報は、検索の鍵として使うことができる他、検索されたメディアデータの情報として、ユーザが使用する検索エンジン11の画面に表示することもできる。

【0068】電子透かしデータ9fは、著作権情報を表すことが多い。その著作権情報を基にその電子透かしデータ9fが埋め込まれたコンテンツが不正コピーされたものか否かをチェックすることができる。不正コピーか否かの情報もデータベース化することができる。不正コピーについては、検索の対象から外したり、警告を発することができる。すなわち、著作権管理ができる。

【0069】図4は、データベースを構築するためのメインルーチンの処理を示すフローチャートである。

【0070】ステップSA1では、初期化処理を行う。例えば、データベース作成の準備を行う。

【0071】ステップSA2では、第1のホームページ（サイト）のアドレスを設定する。第1のホームページは、任意のホームページでよい。具体的には、データベースの管理者がURL（アドレス）を入力する。

【0072】ステップSA3では、上記のアドレスにホームページ（HP）が存在するか否かをチェックする。存在する場合にはステップSA4へ進み、存在しない場合にはステップSA8へ進む。

【0073】ステップSA4では、上記のホームページのコンテンツを取り込む。コンテンツ内に、テキストデータ、楽音データ及び画像データ等の複数種類のデータが含まれているときには、その全てのデータを取り込む。

【0074】ステップSA5では、ホームページのコン

テンツを解析する。まず、コンテンツの分類分けを行い、その後、コンテンツの内容を解析する。このステップの詳細は、後に図5を参照しながら説明する。

【0075】ステップSA6では、解析結果に応じてホームページの情報をデータベースへ登録する。データベースには、コンテンツの種類別に検索ができるように情報を登録する。

【0076】ステップSA7では、取り込んだホームページのコンテンツを削除し、次のホームページの処理に備える。データベースへの登録後は、ホームページのコンテンツが不要になるので、メモリの記憶容量を節約するために、そのコンテンツを削除する。

【0077】ステップSA8では、次のホームページを決定する。1つのホームページには、他のホームページへのリンク先が記述されていることが多い。そのリンク先を、次のホームページとして決定することができる。これにより、自動的に多数のホームページをデータベース化することができる。なお、管理者が次のホームページのURLを入力してもよい。

【0078】その後、ステップSA3に戻り、上記の処理を繰り返す。このデータベース構築装置は、ロボット型であり、定期的に自動巡回し、多数のホームページの登録を行う。

【0079】ホームページは、短周期で更新されたり、新たに作成されたり、削除されたりする。したがって、データベースの内容も短周期で定期的に更新することが好ましい。

【0080】図5は、図4のステップSA5におけるホームページの解析処理の詳細を示すフローチャートである。

【0081】ステップSB1では、HTML形式のテキストデータから単語を抽出する。どのホームページも、通常、テキストデータを含む。

【0082】ステップSB2では、抽出した単語を分析し、登録する語句（単語）を確定する。

【0083】ステップSB3では、テキスト以外のデータをファイル毎に種別化する。データは、種類毎に異なるファイルに収められているので、ファイル毎にデータの種別化を行う。

【0084】ステップSB4では、楽音データ（ウェブデータ又はMIDIデータ）が存在するか否かを判断する。楽音データが存在する場合には、楽音の特質を検出して、楽音のジャンルを登録項目として確定する。このステップの詳細は、後に図6を参照しながら説明する。

【0085】ステップSB5では、ファイルに電子透かしデータが埋め込まれているか否かを判断する。埋め込まれている場合には、電子透かしデータを検出し、著作権情報を登録項目として確定する。

【0086】ステップSB6では、閲覧制限があるか否

かを判断する。例えば、ホームページ内にパスワードの入力画面があるか否かを判断する。その判断結果に応じて、閲覧制限があるか否かの情報を登録項目として確定する。

【0087】ステップSB7では、ホームページ内にリンク先が記述されているか否かを調べる。リンク先が記述されている場合には、そのリンク先を登録項目として確定する。

【0088】ステップSB8では、ホームページの解析日時を取得し、その情報を登録項目とする。ホームページには、高頻度で更新されるものがあるので、解析した日時を登録項目にする必要がある。以上で、ホームページの解析処理を終了し、図4のメインルーチンの処理に戻る。

【0089】図6は、図5のステップSB4におけるMIDIデータの解析処理の詳細を示すフローチャートである。楽音データのうち、MIDIデータについて説明するが、ウェブデータについても同様に処理することができる。

【0090】ステップSC1では、音色データ等の楽音要素を抽出する。例えば、音色データは、MIDIデータ中のプログラムチェンジに相当する。

【0091】ステップSC2では、抽出した楽音要素を基にMIDIデータを分類し、楽音のジャンルを推測する。この際、推測の適正率を付与する。例えば、管楽器の音色又は弦楽器の音色等に分類したり、バイオリンの音色又はピアノの音色等に分類する。

【0092】ステップSC3では、図5のステップSB1で既に抽出された単語データ（テキストデータ）の中に楽音データの内容を推測できるものがある場合には、楽音のジャンルの推測を行い、その推測の適正率と上記の楽音要素から推測したジャンルの適正率とを基に楽音のジャンルを決定する。例えば、推測可能な単語データとして、「バイオリン」や「オーケストラ」等がある。以上で、図5のフローチャートの処理に戻る。

【0093】公知の検索エンジンは、サイト（ホームページ）内のテキストデータを基にデータベースを構築し、そのデータベースを基にサイトの検索を行うものである。テキストデータの検索に適している。

【0094】本実施例によれば、テキストデータはもちろんのこと、サイト内のテキストデータ以外のメディアデータ（例えばウェブデータ、MIDIデータ、静止画データ、動画データ、電子透かしデータ）をも基にしてデータベースを構築する。

【0095】ユーザは、テキストデータを検索できるのみならず、テキストデータ以外の所望のメディアデータをも効率的に検索することができる。この際、メディアデータの種別を指定して、検索することができる。例えば、MIDIデータのみを検索することができる。

【0096】検索する際の鍵（キー）は、テキストデー

タでもよいし、テキストデータ以外のメディアデータでもよい。例えば、「バイオリン」を鍵としてMIDIデータを検索したり、所定の楽譜（フレーズ）を鍵としてMIDIデータを検索することもできる。

【0097】コンテンツ内に著作権情報が電子透かしデータとして埋め込まれている場合、その著作権情報に基づいてそのコンテンツが不正コピーされたものか否かをチェックし、不正コピーか否かの情報もデータベース化することができる。不正コピーについては、検索の対象から外したり、警告を発することができる。

【0098】なお、本実施例は、インターネットに限定されず、その他のネットワークに適用することもできる。

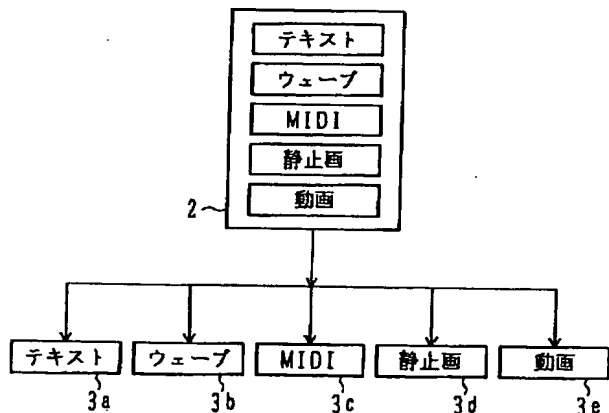
【0099】以上実施例に沿って本発明を説明したが、本発明はこれらに制限されるものではない。例えば、種々の変更、改良、組み合わせ等が可能なことは当業者に自明であろう。

【0100】

【発明の効果】以上説明したように、本発明によれば、データベースには、テキストデータ以外のメディアデータに関連する鍵とそのデータが存在するネットワーク上のサイトとが対応付けて登録される。ユーザは、テキストデータ以外のメディアデータ（例えば、楽音のウェブデータ、MIDIデータ、静止画データ、動画データ又は電子透かしデータ）を効率的に検索することができ

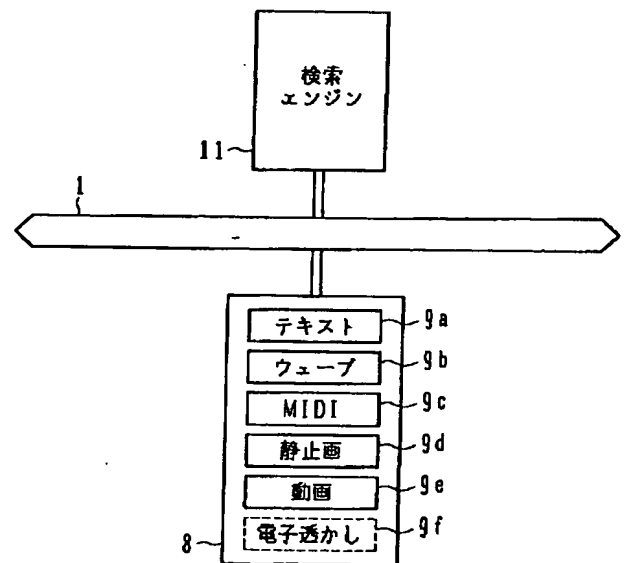
【図2】

データ分類



【図3】

データベース検索



る。

【図面の簡単な説明】

【図1】 本発明の実施例によるデータベース構築装置の構成を示すブロック図である。

【図2】 メディアデータの分類方法を示すブロック図である。

【図3】 検索エンジンを用いて検索を行うためのネットワークの構成図である。

【図4】 データベースの構築方法を示すフローチャートである。

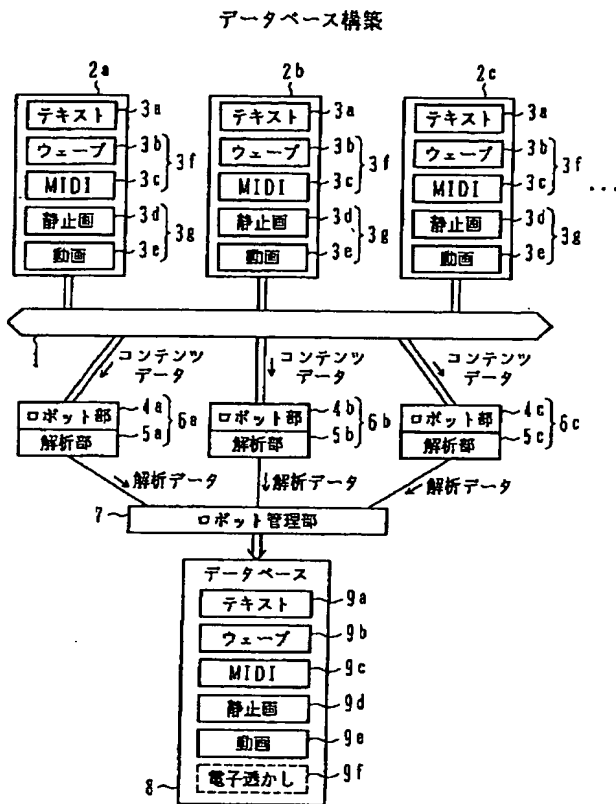
【図5】 図4のステップSA5におけるホームページの解析処理の詳細を示すフローチャートである。

【図6】 図5のステップSB4におけるMIDIデータの解析処理の詳細を示すフローチャートである。

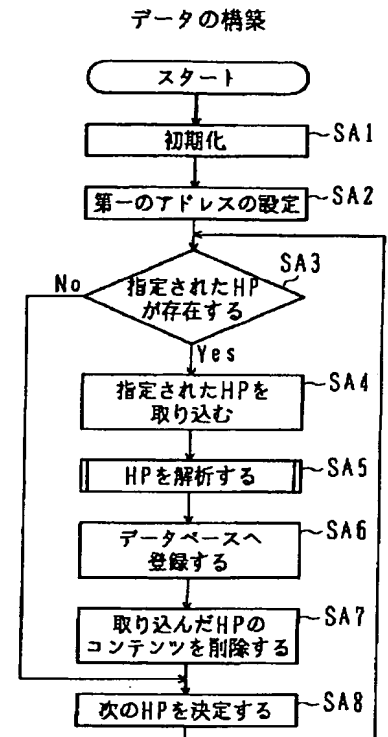
【符号の説明】

1 インターネット回線、 2 サイト、 3 a, 9 a テキストデータ、 3 b, 9 b ウェブデータ、 3 c, 9 c MIDIデータ、 3 d, 9 d 静止画データ、 3 e, 9 e 動画データ、 3 f 楽音データ、 3 g 画像データ、 4 巡回ロボット部、 5 コンテンツ解析部、 6 巡回及び解析部、 7 ロボット管理部、 8 データベース、 9 f 電子透かしデータ、 11 検索エンジン

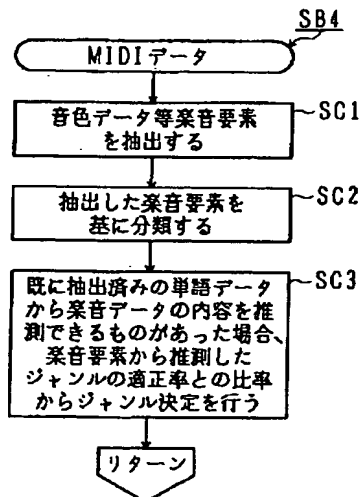
【図1】



【図4】



【図6】



【図5】

